IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 5, 2018, accepted October 12, 2018, date of publication October 22, 2018,
date of current version November 30, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2876880

Multiple Graph Regularized Concept Factorization With Adaptive Weights

ZHENQIU SHU[1], XIAO-JUN WU[2], PU HUANG[3], HONGHUI FAN[1], ZHEN LIU[2], AND FEIYUE YE[1]

[1]School of Computer Engineering, Jiangsu University of Technology, Changzhou 231001, China
[2]School of IoT Engineering, Jiangnan University, Wuxi 214122, China
[3]Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Corresponding author: Feiyue Ye (yfy_jsut@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61472166, Grant 61603159, Grant 61672265, and Grant U1831127, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20160293, in part by the China Postdoctoral Science Foundation under Grant 2017M611695, and in part by the Jiangsu Province Postdoctoral Science Foundation under Grant 1701094B.

ABSTRACT Many traditional concept factorization methods employ single graph to approximate the manifold structure of data. Therefore, they cannot capture the underlying geometric structure hidden in data effectively. In this paper, we propose a novel method, called Multiple graph regularized Concept Factorization with Adaptive Weights (MCFAWs), for data representation. It exploits the intrinsic geometric manifold of the data by the linear combination of multiple graphs with parameter free. Therefore, our proposed MCFAW method can be applied to many real problems. Besides, an efficient optimization algorithm is presented to solve the proposed model. Some experimental results on the benchmarks show that the proposed MCFAW method outperforms the state-of-the-art methods.

INDEX TERMS Concept factorization; manifold; graph; parameter free; data representation

I. INTRODUCTION

Over the past few years, representation learning of the high dimensional data is a fundamental topic in pattern recognition and machine learning communities. Data representation aims to explore the semantic information of the high dimensional data using the low dimensional features, which is an effective way to overcome the curse of dimensionality. Recently, there have been several attempts to improve the performances of data representation techniques [1]–[5].

Many studies have shown that matrix factorization methods can provide a powerful representation ability in real applications, such as data clustering, object tracking and image classification [6]–[8]. Among the matrix factorization methods, Nonnegative Matrix Factorization (NMF) [9] requires both basis matrix and coefficient matrix to be strictly nonnegative, and thus only allows additive, not subtractive, combination of the samples. Therefore, it is a parts-based representation method. This strict constraint, however, leads to the fact that NMF cannot be applied to many real problems due to noise or outlier. To solve this problem, the Concept Factorization (CF) [10] method was proposed for document clustering. The major advantages of CF over NMF is that CF is not only performed on the high dimensional data mixed with negative elements, but also can be easily extended to its kernel version. In order to make full use of the label information among the data, many semi-supervised and supervised methods have developed in the past few years [11]–[16].

Recently, Liu *et al.* [11] proposed a semi-supervised concept factorization method by adding a hard label constraint. Hua *et al.* [12] introduced a supervised method, call discriminative concept factorization (DCF) for data representation. Li *et al.* [13] proposed a semi-supervised discriminative concept factorization method that adopts the limited label among the data as a discriminative constraint. In order to explore the latent manifold of data, Cai *et al.* [17] proposed the Local Consistent Concept Factorization (LCCF) method using single graph regularizer. Shu *et al.* [18] further introduced a local learning regularized CF method, which uses each sample's neighbors to exploit both the local manifold structure and the discriminative structure of data. A locality-constrained concept factorization (LCF) algorithm [19] was proposed by incorporating a locality constraint into model of CF. Inspired by deep learning, the multilayer concept factorization (MCF) method [20] was proposed for data representation. MCF utilizes the multilayer structure to learn the representation of high dimensional data. To take advantage

of the manifold structures of both data space and feature space, they further proposed a dual graph regularized MCF method. In order to adaptively determine the neighborhood number of each sample in graph model, a variant of CF, called CF with adaptive neighbors (CFANs) [21], was proposed by imposing the adaptive neighbors constraint. Therefore, it can effectively extract the representation space that preserves geometrical structure of the data. The methods mentioned above adopt single graph model to approximate the manifold structure of data. However, they cannot effectively discover the underlying structure hidden in data in this way.

In this paper, we propose a novel method, called Multiple graph Concept Factorization with Adaptive Weights (MCFAW), to better explore the intrinsic manifold structure of data. Our proposed method employs a linear combination of multiple graphs to construct a regularizer. Moreover, the weight set of these graphs is determined adaptively without introducing additional parameter. Therefore, our proposed MCFAW method is easily applied to a wide range of practical problems. Experimental results have shown that the proposed method is superior to other methods on some benchmarks.

The remainder of this paper is organized as follows. We represent an overview of both NMF and CF in Section 2. We describe our proposed MCFAW method and its optimization scheme in Section 3. Experimental results are shown in Section 4. Finally, we draw the conclusions in Section 5.

II. RELATED WORKS

A. NMF

The goal of NMF is to represent the input matrix $X \in R^{m\times n}$ as the product of two low-rank nonnegative matrices $U \in R^{m\times k}$ and $V \in R^{n\times k}$. Therefore, the standard NMF minimizes the objective function as follows:

$$O = \left\|X - UV^T\right\|_F^2 \quad s.t.\ U > 0, \quad V > 0 \tag{1}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. It is obvious to see that it is unrealistic to seek its global optimization solution due to the non-convexity of Eq. (1). Fortunately, its local optimization minimum can be achieved using an alternative optimization algorithm [9]. Therefore, the updating rules of model (1) can be presented as follows:

$$u_{ij}^{t+1} \leftarrow u_{ij}^t \frac{(XV)_{ij}}{(UV^TV)_{ij}} \tag{2}$$

$$v_{ij}^{t+1} \leftarrow v_{ij}^t \frac{(X^TU)_{ij}}{(VU^TU)_{ij}} \tag{3}$$

B. CF

CF is an important variation of NMF. In CF, each underlying concept can be characterized by a linear combination of all samples $u_j = \sum_i h_{ij}x_i$, where h_{ij} denotes a nonnegative association weight. Therefore, CF aims to seek the following approximation

$$X \approx XHV^T \tag{4}$$

by minimizing the following problem

$$O = \left\|X - XHV^T\right\|_F^2 \quad s.t.\ H > 0, \quad V > 0 \tag{5}$$

Similarly, Eq. (5) can be solved by the similar optimization scheme. Therefore, the updating rules of CF are given as follows:

$$h_{ij}^{t+1} \leftarrow h_{ij}^t \frac{(KV)_{ij}}{(HV^TV)_{ij}} \tag{6}$$

$$v_{ij}^{t+1} \leftarrow v_{ij}^t \frac{(KH)_{ij}}{(VH^TKH)_{ij}} \tag{7}$$

where $K = X^TX$. From Eq. (6) and (7), it can be observed that CF is easily kernelized while dealing with nonlinear problems.

III. OUR PROPOSED METHOD

A. MOTIVATION

It is often acknowledged that the manifold learning theory plays an important role in data representation. In traditional CF methods, a single nearest neighbor graph is used to model the manifold structure hidden in data. However, it is difficult to determine the optimal number of the nearest neighbors. Therefore, we propose a novel method, called Multiple graph regularized Concept Factorization with Adaptive Weights (MCFAW), to explore the geometric structure of the high dimensional data. Our MCFAW can learn an optimal weight set for all graphs automatically without introducing additional parameter. Therefore, our proposed method is easily to be applied to real problems. We will introduce the model of our proposed method and its optimization scheme in details in the following subsections.

B. THE MODEL OF THE PROPOSED MCFAW METHOD

Our proposed MCFAW method explores the intrinsic manifold that are approximated by the linear combination of multiple Laplacian graphs. Therefore, the model of MCFAW can be formulated as the following minimization problem:

$$O = \left\|X - XHV^T\right\|_F^2 + \lambda \sum_{i=1}^{q} \alpha_i Tr(V^TL_iV)$$
$$s.t.\ H > 0, V > 0, \alpha_i = 1/(2\sqrt{Tr(V^TL_iV)}) \tag{8}$$

where λ denotes a nonnegative tradeoff parameter, and α_i is the weight of the i-th Laplacian graph. The first term denotes the reconstruction term, and the second term stands for our proposed multiple graph regularizer.

C. ALGORITHM FOR SOLVING THE PROPOSED MODEL

It is clear that the model of our proposed MCFAW method is not-convex in H and V together. Fortunately, we can obtain a sub-optimal local minimum solution using the multiplicative updating algorithm proposed in [9].

Then our proposed model (8) can be rewritten as follows:

$$O = \left\|X - XHV^T\right\|_F^2 + \lambda \sum_{i=1}^{q} \alpha_i Tr(V^TL_iV)$$

$$= (X - XHV^T)^T(X - XHV^T) + \lambda \sum_{i=1}^{q} \alpha_i Tr(V^T L_i V)$$
$$= (I - HV^T)^T K(I - HV^T) + \lambda \sum_{i=1}^{q} \alpha_i Tr(V^T L_i V)$$
$$= tr(K) - 2Tr(VH^T K) + Tr(VH^T KHV^T) + \lambda \sum_{i=1}^{q} \alpha_i Tr(V^T L_i V) \quad (9)$$

where $K = X^T X$. Therefore, the problem (9) can be solved by iterating the following three steps:

(a) Optimal Solution to H

Let ψ_{ik} be the Lagrange multiplier for constraint $h_{ik} \geq 0$. Defining $\Psi = [\psi_{ik}]$, the Lagrangian function $\mathcal{L}(H)$ is given by

$$\mathcal{L}(H) = -2Tr(VH^T K) + Tr(VH^T KHV^T) + Tr(\Psi H^T) \quad (10)$$

By taking the derivative of Eq. (10) with respect to H, we have

$$\frac{\partial \mathcal{L}(H)}{\partial H} = -2KV + 2KHV^T V + \Psi \quad (11)$$

According to Karush-Kuhn-Tucker conditions $\psi_{ik} h_{ik} = 0$, we have

$$-(KV)_{ik} u_{ik} + (KHV^T V)_{ik} u_{ik} = 0 \quad (12)$$

Thus, we derive the updating rule of H as follows:

$$h_{ij}^{t+1} \leftarrow h_{ij}^{t} \frac{(KV)_{ij}}{(KHV^T V)_{ij}} \quad (13)$$

(b) Optimal Solution to V

Let ϕ_{ik} be the Lagrange multiplier for constraint $v_{jk} \geq 0$. Defining $\Phi = [\phi_{ik}]$, the Lagrangian function $\mathcal{L}(V)$ is defined by

$$\mathcal{L}(V) = -2Tr(VH^T K) + Tr(VH^T KHV^T) + \lambda \sum_{i=1}^{q} \alpha_i Tr(V^T L_i V) + Tr(\Phi V^T) \quad (14)$$

By taking the derivative of Eq. (14) with respect to V, we have

$$\frac{\partial \mathcal{L}(V)}{\partial V} = -2KU + 2VH^T KH + 2\lambda \sum_{i=1}^{q} \alpha_i L_i V + \Phi \quad (15)$$

According to Karush-Kuhn-Tucker conditions $\phi_{jk} v_{jk} = 0$, we get the following equation

$$-(KH)_{jk} v_{jk} + (VH^T KH)_{jk} v_{jk} + (\lambda \sum_{i=1}^{q} \alpha_i L_i V)_{jk} v_{jk} = 0 \quad (16)$$

Therefore, Eq. (16) can lead to the following updating rule:

$$v_{ij}^{t+1} \leftarrow v_{ij}^{t} \frac{(KH + \lambda \sum_{i=1}^{q} \alpha_i W_i V)_{ik}}{(VH^T KH + \lambda \sum_{i=1}^{q} \alpha_i D_i V)_{ik}} \quad (17)$$

(c) Optimal Solution to α_i

Our proposed method can automatically assign the optimal weights to all Laplacian graphs. In other words, the larger weight value is assigned to a better graph, and vice versa. Here, the weight α_i of the i-th Laplacian graph can be set as

$$\alpha_i = 1/(2\sqrt{Tr(V^T L_i V)}) \quad (18)$$

It can be seen that the weight α_i can be determined by the variable V. Obviously, the weight α_i can be iteratively updated according to Eq. (18).

D. THE PROPOSED MCFAW ALGORITHM

According to the above description, the flowchart of our proposed MCFAW method is summarized as follows.

Algorithm 1 Our Proposed MCFAW Method

Input: A data set of m samples $X = [x_1, x_2, \cdots, x_m]$, iteration number t, regularization parameter λ, Laplacian graph number q.

Step 1: Initialize the weight factor $\alpha_i = \frac{1}{q}$;

Step 2: Construct q nearest neighbour graphs $(W_1, W_2, ..., W_q)$;

Step 3: Calculate the Laplacian matrices $(L_1, L_2, ..., L_q)$ and the diagonal matrices $(D_1, D_2, ..., D_q)$, respectively;

Step 4: For $i = 1, \ldots, t$ do

(a) Update the basis matrix H by Eq.(13);

(b) Update the coefficient matrix V by Eq.(17);

(c) Update the weight α_i by Eq.(18);

End for

Output: basis matrix H and coefficient matrix V.

IV. EXPERIMENTAL ANALYSIS

In this section, we carry out some experiments to investigate the proposed MCFAW method for image clustering. To show the superiority of the proposed method, we compare it with five other related methods, such as K-means, PCA, NMF, CF and LCCF, on COIL20, MNIST and PIE datasets. Two measures including Accuracy (AC) and Normalized Mutual Information (NMI) are adopted to evaluate all methods for clustering analysis.

A. COIL20 IMAGE DATASET

The COIL20 dataset includes 1440 image samples from 20 objects. Each object was taken 72 images from different angles. For simplicity, all sample images were manually cropped to 32×32, and thus can be represented as a matrix of size 1024×1440. Some samples from the COIL20 image database are shown in Fig.1.

FIGURE 1. Sample images from the COIL20 dataset

TABLE 1. Experimental results on the COIL20 database.

AC						
N	K-means	PCA	NMF	CF	LCCF	MCFAW
8	0.793	0.781	0.787	0.757	0.765	0.855
10	0.704	0.664	0.668	0.618	0.673	0.728
12	0.664	0.680	0.691	0.6388	0.644	0.753
14	0.668	0.689	0.684	0.570	0.647	0.737
16	0.659	0.696	0.672	0.559	0.656	0.727
18	0.594	0.645	0.618	0.554	0.612	0.661
20	0.616	0.612	0.638	0.546	0.607	0.693
avg	0.671	0.681	0.680	0.606	0.658	0.736
NMI						
N	K-means	PCA	NMF	CF	LCCF	MCFAW
8	0.785	0.774	0.770	0.750	0.762	0.848
10	0.732	0.718	0.711	0.670	0.728	0.769
12	0.727	0.7316	0.731	0.691	0.726	0.797
14	0.754	0.757	0.741	0.664	0.734	0.798
16	0.745	0.762	0.754	0.646	0738	0.793
18	0.718	0.743	0.712	0.659	0.709	0.763
20	0.731	0.734	0.732	0.657	0.723	0.782
avg	0.742	0.746	0.736	0.677	0.731	0.793

In this experiment, we randomly selected N categories samples as the subset to evaluate all methods. Here, every method was run 10 times on different randomly chosen clusters and its average performance was reported as the final result. For our proposed method, we picked out 5, 7 and 9 neighbors for each sample to construct three graph models, respectively. The clustering results of all methods on COIL database are shown in Table 1. It is clear to see that our proposed MCFAW method outperforms consistently compared with other methods. This is because MCFAW uses the linear combination of multiple graphs to explore the manifold structure of the data. Therefore, it can capture the intrinsic geometric structure of data effectively. Moreover, our proposed MCFAW method is parameter free.

FIGURE 2. Sample images from the MNIST dataset

TABLE 2. Experimental results on the MNIST database.

AC						
N	K-means	PCA	NMF	CF	LCCF	MCFAW
4	0.708	0.792	0.670	0.642	0.647	0.767
5	0.544	0.535	0.5728	0.580	0.587	0.584
6	0.654	0.680	0.621	0.640	0.644	0.673
7	0.564	0.605	0.546	0.578	0.582	0.641
8	0.604	0.583	0.552	0.596	0.604	0.623
9	0.566	0.571	0.558	0.535	0.545	0.578
10	0.570	0.550	0.5276	0.541	0.530	0.551
avg	0.601	0.617	0.578	0.587	0.591	0.631
NMI						
N	K-means	PCA	NMF	CF	LCCF	MCFAW
4	0.555	0.582	0.496	0.484	0.498	0.610
5	0.451	0.450	0.428	0.436	0.441	0.495
6	0.577	0.565	0.504	0.543	0.553	0.588
7	0.507	0.517	0.484	0.505	0.519	0.576
8	0.558	0.533	0.501	0.536	0.537	0.565
9	0.545	0.538	0.522	0.511	0.512	0.546
10	0.553	0.521	0.499	0.508	0.505	0.528
avg	0.535	0.529	0.491	0.503	0.509	0.558

FIGURE 3. Sample images from the PIE database. (a) Results on the COIL20 dataset. (b) Results on the MNIST dataset. (c) Results on the PIE dataset.

B. MNIST HANDWRITTEN DIGITS DATASET

The MNIST handwritten digits dataset contains a total of 10000 image samples. For simplicity, we randomly chose 500 sample images as the experimental subset data. Each handwritten digital sample was normalized to 28×28 gray scale image. Fig. 2 shows some handwritten digits images from the MNIST dataset.

This experiment adopted the same scheme as the previous experiment. Similarly, we also selected different neighbor-

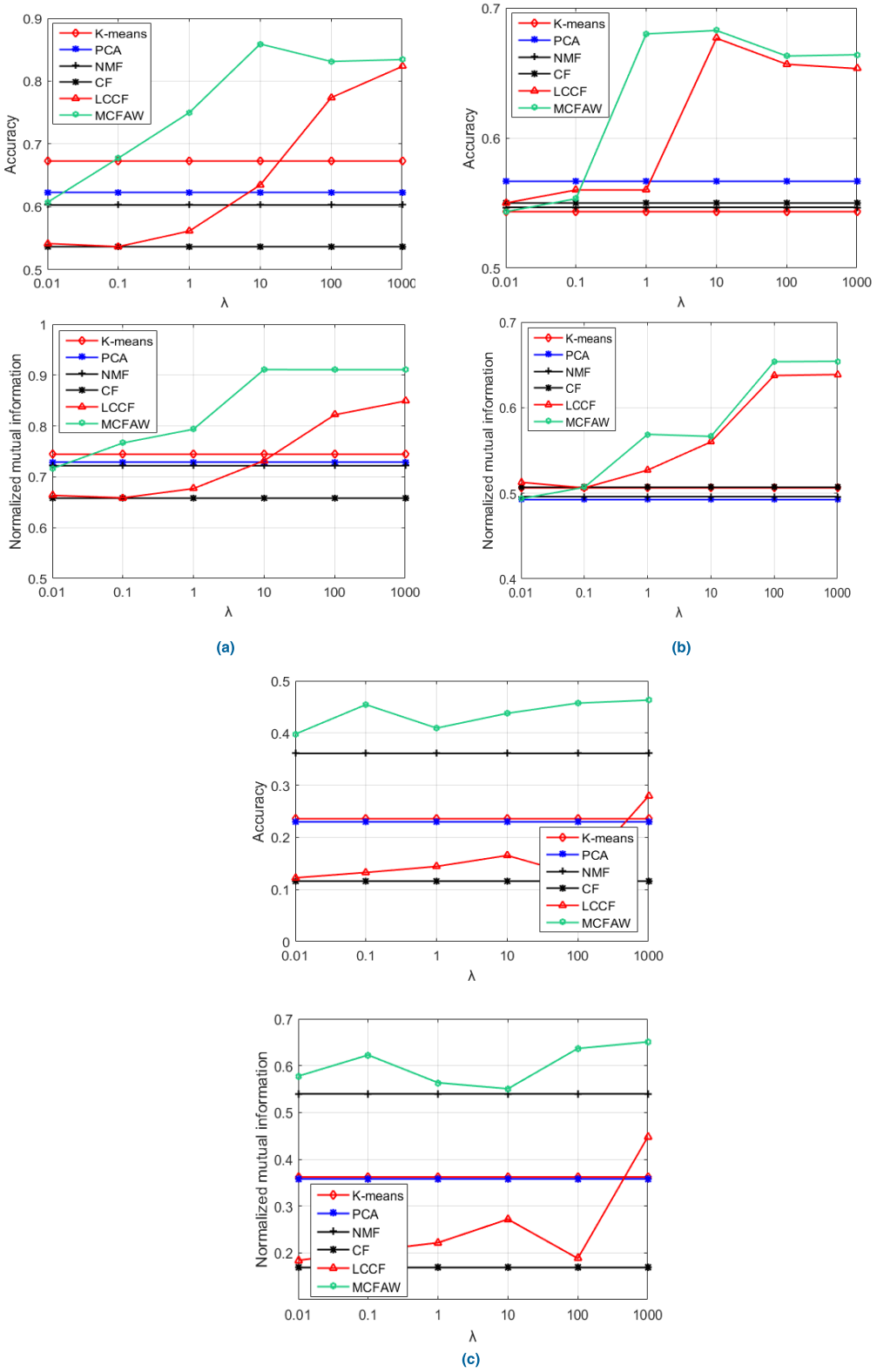

FIGURE 4. Clustering performance versus the parameter λ. (a) Results on the COIL20 dataset. (b) Results on the MNIST dataset. (c) Results on the PIE dataset.

hood samples for each sample to construct three graph models, respectively. Table 2 shows the performance of all methods on MNIST database. It can be observed that our proposed MCFAW method achieves the best average performance in all methods. The main reason is that the learned coefficient matrix can encode the geometry manifold structure

TABLE 3. Experimental results on the PIE database.

AC						
N	K-means	PCA	NMF	CF	LCCF	MCFAW
20	0.206	0.199	0.322	0.185	0.186	0.325
30	0.211	0.197	0.307	0.177	0.178	0.329
40	0.199	0.195	0.319	0.153	0.170	0.334
50	0.190	0.181	0.296	0.146	0.165	0.338
60	0.191	0.183	0.295	0.149	0.169	0.336
68	0.179	0.181	0.290	0.146	0.156	0.327
avg	0.196	0.189	0.305	0.159	0.170	0.332
NMI						
N	K-means	PCA	NMF	CF	LCCF	MCFAW
20	0.270	0.262	0.438	0.134	0.158	0.455
30	0.336	0.321	0.488	0.140	0.173	0.503
40	0.356	0.362	0.520	0.179	0.215	0.544
50	0.377	0.371	0.516	0.182	0.217	0.556
60	0.391	0.394	0.532	0.194	0.234	0.577
68	0.402	0.405	0.542	0.212	0.248	0.579
avg	0.355	0.353	0.506	0.174	0.208	0.536

information of the data effectively using multiple graph regularizer without additional parameter. Therefore, it is easily applied to many practical problems.

C. PIE FACE DATASET

The PIE face database includes 41,368 face images of 68 individuals, which were taken at different poses, illumination and expressions. Here, 46 images per class were selected to evaluate the proposed method. Each face image was converted as a 1024 dimensional vector. Fig. 3 shows some sample images from the PIE database.

Similarly, we also randomly selected N categories samples as the experimental subset. All methods were run ten times on each value of N and their average performances were reported. Meanwhile, 5, 7 and 9 neighborhoods of each sample were also used to construct three nearest neighborhood graphs. The experimental results of all methods are shown in Table 3. We can see that LCCF can outperform CF all configuration of cluster number N. The main reason is that LCCF considers the geometric manifold structure of the data using graph regularizer. Besides, it can be seen that our proposed MCFAW method can achieve a significant improvement in comparison with LCCF in terms of AC and NMI. This is because that MCFAW can automatically select the best neighborhood graph without additional parameter. Therefore, the proposed method can learn a better representation for the high dimensional data than other methods.

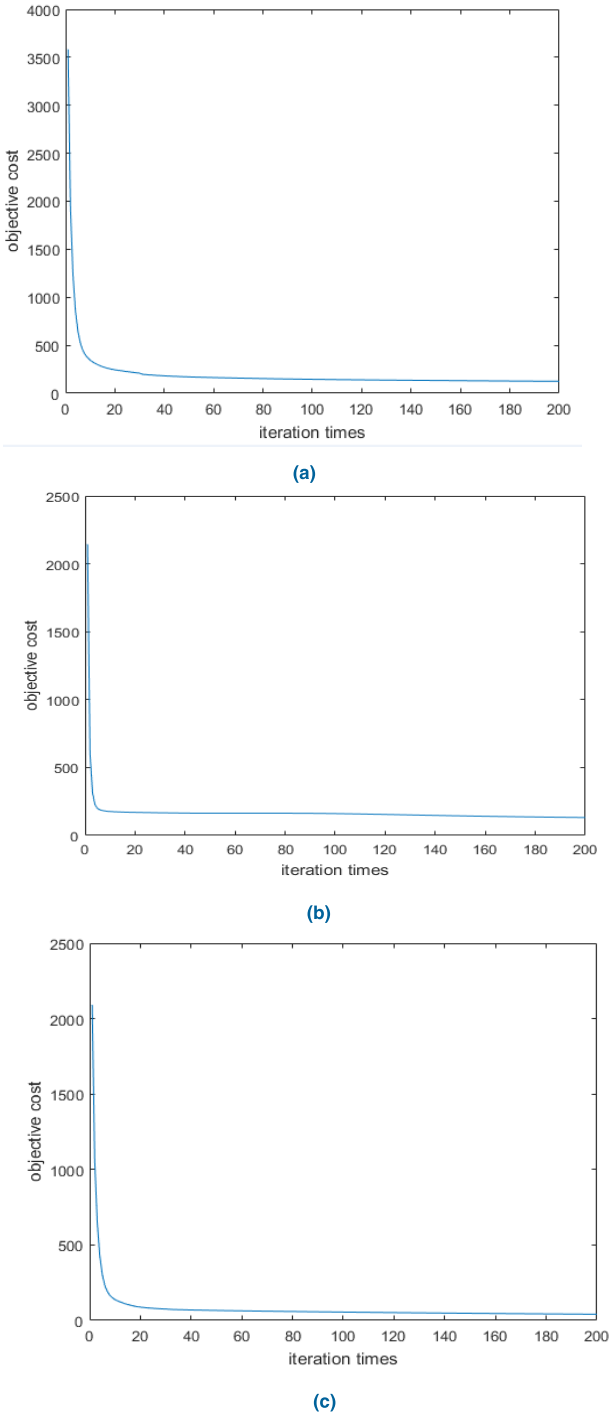

FIGURE 5. Convergence curves of our proposed method. (a) COIL20, (b) MNIST, (c) PIE.

D. PARAMETER SELECTION

A regularization parameter λ need to be set in the models of both LCCF and MCFAW. In this subsection, some experiments are conducted to evaluate the proposed MCFAW method with varied parameter values.

Here, we randomly chose samples in 14, 5, and 30 categories from COIL20, MNIST and PIE datasets, respectively.

The parameter λ was set by the grids {0.01, 0.1, 1, 10, 100, 1000}. FIGURE 4. shows the performances of all methods with varied parameter λ. It can be seen that the proposed MCFAW method achieves a stable performance in a large range of the values of the parameter λ.

E. CONVERGENCE ANALYSIS

In this subsection, we carry out some experiments to clearly show the convergence of our proposed MCFAW method. Similarly, 14, 5, and 30 categories samples were chosen from COIL20, MNIST and PIE datasets as the data subset. The curves about the objective cost on three benchmarks are given in Fig. 5. From the results in Fig. 5, we can see that our proposed MCFAW method can converge within 40 iterations.

V. CONCLUSION

In this paper, we propose a novel method, called Multiple graph regularized Concept Factorization with Adaptive Weights (MCFAW), which uses a linear combination of multiple Laplacian graphs to approximate the intrinsic manifold structure of data. Moreover, the weights of these graphs can be learned automatically without additional parameter. Therefore, our proposed MCFAW method could be easily applied to many real problems. Extensive experiments are conducted on three benchmark datasets, and the results have shown that our proposed method achieves better clustering performance than the state-of-the-art methods.

REFERENCES

[1] X. Shen, W. Liu, I. W. Tsang, Q.-S. Sun, and Y.-S. Ong, "Multilabel prediction via cross-view search," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4324–4338, Sep. 2018.
[2] Z. Shu, X. Wu, and H. Fan, "Parameter-less auto-weighted multiple graph regularized nonnegative matrix factorization for data representation," *Knowl.-Based Syst.*, vol. 131, pp. 105–112, Sep. 2017.
[3] X. Shen, F. Shen, Q.-S. Sun, Y. Yang, Y.-H. Yuan, and H. T. Shen, "Semi-paired discrete hashing: Learning latent hash codes for semi-paired cross-view retrieval," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4275–4288, Dec. 2016.
[4] P. Huang, T. Li, Z. Shu, G. Gao, G. Yang, and C. Qian, "Locality-regularized linear regression discriminant analysis for feature extraction," *Inf. Sci.*, vol. 429, pp. 164–176, Mar. 2018.
[5] Z. Shu, J. Zhou, P. Huang, X. Yu, Z. Yang, and C. Zhao, "Local and global regularized sparse coding for data representation," *Neurocomputing*, vol. 175, pp. 188–197, Jan. 2016.
[6] R. Hong, Y. Yang, M. Wang, and X.-S. Hua, "Learning visual semantic relationships for efficient visual retrieval," *IEEE Trans. Big Data*, vol. 1, no. 4, pp. 152–161, Dec. 2015.
[7] R. Hong, L. Lei, J. Cai, D. Tao, M. Wang, and Q. Tian, "Coherent semantic-visual indexing for large-scale image retrieval in the cloud," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4128–4138, Sep. 2017.
[8] Z. Li, J. Liu, and H. Lu, "Structure preserving non-negative matrix factorization for dimensionality reduction," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1175–1189, 2013.
[9] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, Oct. 1999.
[10] W. Xu and Y. Gong, "Document clustering by concept factorization," in *Proc. ACM SIGIR*, 2004, pp. 202–209.
[11] H. Liu, G. Yang, Z. Wu, and D. Cai, "Constrained concept factorization for image representation," *IEEE Trans. Cybern.*, vol. 44, no. 7, pp. 1214–1224, Jul. 2014.
[12] W. Hua and X. He, "Discriminative concept factorization for data representation," *Neurocomputing*, vol. 74, no. 18, pp. 3800–3807, 2011.
[13] H. Li, J. Zhang, J. Hu, C. Zhang, and J. Liu, "Graph-based discriminative concept factorization for data representation," *Knowl.-Based Syst.*, vol. 118, pp. 70–79, Feb. 2017.
[14] M. Lu, X.-J. Zhao, L. Zhang, and F.-Z. Li, "Semi-supervised concept factorization for document clustering," *Inf. Sci.*, vol. 331, pp. 86–98, Feb. 2016.
[15] Y. He, H. Lu, L. Huang, and S. Xie, "Pairwise constrained concept factorization for data representation," *Neural Netw.*, vol. 52, pp. 1–17, Apr. 2014.
[16] M. Lu, L. Zhang, X.-J. Zhao, and F.-Z. Li, "Constrained neighborhood preserving concept factorization for data representation," *Knowl.-Based Syst.*, vol. 102, pp. 127–139, Jun. 2016.
[17] D. Cai, X. He, and J. Han, "Locally consistent concept factorization for document clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 6, pp. 902–913, Jun. 2011.
[18] Z. Shu, C. Zhao, and P. Huang, "Local regularization concept factorization and its semi-supervised extension for image representation," *Neurocomputing*, vol. 158, pp. 1–12, Jun. 2015.
[19] H. Liu, Z. Yang, and Z. Wu, "Locality-constrained concept factorization," in *Proc. 22nd Int. Joint Conf. Artif. Intell. (IJCAI)*, Barcelona, Spain, 2011, pp. 1378–1383.
[20] X. Li, X. Shen, Z. Shu, Q. Ye, and C. Zhao, "Graph regularized multilayer concept factorization for data representation," *Neurocomputing*, vol. 238, pp. 139–151, May 2017.
[21] X. Pei, C. Chen, and W. Gong, "Concept factorization with adaptive neighbors for document clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 2, pp. 343–352, Feb. 2016.

ZHENQIU SHU received the Ph.D. degree in computer applications from the Nanjing University of Science and Technology. In 2015, he joined the School of Computer Engineering, Jiangsu University of Technology. He is currently a Post-Doctoral Fellow with the School of IoT Engineering, Jiangnan University. His research interests include machine learning, data mining, and pattern recognition.

XIAO-JUN WU received the B.S. degree in mathematics from Nanjing Normal University, Nanjing, China, in 1991, and the M.S. degree and the Ph.D. degree in pattern recognition and intelligent system from the Nanjing University of Science and Technology, Nanjing, China, in 1996 and 2002, respectively. He was a fellow of the International Institute for Software Technology, United Nations University, from 1999 to 2000. From 1996 to 2006, he taught at the School of Electronics and Information, Jiangsu University of Science and Technology, where he was an exceptionally promoted Professor. From 2003 to 2004, he was a Visiting Postdoctoral Researcher with the Centre for Vision, Speech, and Signal Processing, University of Surrey, U.K., under the supervision of Prof. J. Kittler. In 2006, he joined the School of Information Engineering (School of IoT Engineering), Jiangnan University, where he is currently a Distinguished Professor of level two. He has published over 200 papers in his fields of research, among which 60 papers have indexed by SCI and over 100 papers have been indexed by EI. His current research interests are pattern recognition, computer vision, fuzzy systems, neural networks, and intelligent systems. He received the Most Outstanding Postgraduate Award from the Nanjing University of Science and Technology. He received different awards, including the International Award, the National Award, and the Provincial Award, for his research achievements.

PU HUANG received the B.S. and M.S. degrees in computer applications from Yangzhou University, China, in 2007 and 2010, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from the Nanjing University of Science and Technology, China, in 2014. He joined the Nanjing University of Posts and Telecommunications in 2014. His research interests include pattern recognition, computer vision, and machine learning.

HONGHUI FAN received the Ph.D. degree in computer application technology from Yamagata University, Japan, in 2011. He is currently an Associate Professor with the School of Computer Engineering, Jiangsu University of Technology. His research interests include image process and computer vision.

ZHEN LIU is currently pursuing the Ph.D. degree with Jiangnan University, Wuxi. His research interests include face processing and recognition, computer vision, and syntactic pattern recognition.

FEIYUE YE received the Ph.D. degree from the School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, in 2006. He is currently a Professor with the School of Computer Engineering, Jiangsu University of Technology. His research interests include data mining and machine learning.